**Traffic Accidents Analysis Dashboard Report**  
*Your Name: Jas Shah*  
*SBU ID: 116633222*

**Aim**

The primary aim of this project is to develop an interactive dashboard that facilitates a deeper understanding of traffic accident data. By integrating advanced techniques such as Principal Component Analysis (PCA) and K-Means clustering, along with dynamic visualizations (including scree plots, elbow plots, biplots, and scatterplot matrices), the dashboard allows users to uncover latent patterns and relationships within the dataset. This, in turn, provides valuable insights into the factors that might influence accident severity and frequency.

**Data Sources**

The dataset used in this project is provided via a CSV file named *traffic\_accidents\_dict new.csv*.

Kaggle Link: https://www.kaggle.com/datasets/willianoliveiragibin/key-factors-traffic-accidents

**Understanding the Attributes**

In preparing the data for analysis, the dashboard script selects only numerical columns to ensure the applicability of PCA and K-Means clustering. Some common attributes in a traffic accidents dataset that might be included are:

1. **Accident Severity:** Numerical score or rating that may reflect the impact of the accident.
2. **Number of Vehicles Involved:** Indicates how many vehicles were involved in the accident.
3. **Speed:** Recorded speeds at the time of the accident.
4. **Time of Day:** Encoded as a numerical value to capture the temporal dimension.
5. **Weather Conditions:** Numerical representations (if available) to quantify visibility or road conditions.
6. **Road Type or Location Indicators:** Additional metrics that can be quantified to provide context.

During preprocessing, rows with missing values are removed and non-numerical columns are dropped or converted as needed. This ensures that the subsequent PCA and clustering algorithms work with a clean, uniform dataset.

**Why is This Data Interesting?**

Traffic accidents are a critical public safety issue, and understanding their underlying patterns can lead to actionable insights. Key points include:

* **Identifying High-Risk Scenarios:** By using PCA to reduce dimensionality, the dashboard highlights the most significant variables, which may correlate with accident severity.
* **Clustering for Patterns:** K-Means clustering can reveal groups of accidents that share common characteristics, potentially identifying high-risk clusters or time periods.
* **Visual Insights:** Interactive visualizations, such as the scree plot and elbow plot, allow users to dynamically select the intrinsic dimensions and optimal cluster counts, providing an exploratory approach to data analysis.
* **Actionable Strategies:** The insights gleaned from the dashboard can support decision-makers in developing targeted safety interventions, improving road designs, or enhancing driver education programs.

**Implementation**

**Overview**

The dashboard was implemented using Dash (a Python web application framework) alongside Plotly Express and Graph Objects for dynamic and interactive visualizations. The project is organized into several tabs, each offering a unique view into the data analysis:

1. **PCA & K-Means Tab:**
   * **Scree Plot:** Displays the eigenvalues of the PCA, allowing users to select the number of dimensions by clicking on a bar.
   * **Elbow Plot:** Visualizes the K-Means inertia values for different cluster counts, with user interaction to choose the optimal number of clusters.
2. **Biplot Tab:**
   * Features a PCA biplot that shows data points (colored by cluster assignment) alongside feature vectors. A slider lets users control the number of feature vectors displayed, enhancing the exploration of feature directions.
3. **Top PCA Features Tab:**
   * Provides a table summarizing the top features based on PCA loadings. This helps identify which variables contribute most significantly to the principal components.
4. **Scatterplot Matrix Tab:**
   * Offers a comprehensive view of the relationships between the top features through a scatterplot matrix. Points are colored based on cluster membership, and jitter is applied to categorical variables for improved clarity.

**Visuals and Interactivity**

* **Scree Plot:** Users can click on the bars representing each principal component to dynamically set the intrinsic dimensionality. A cumulative variance line further helps in understanding how much variance is captured.
* **Elbow Plot:** By clicking on the points of the elbow plot, users can select the optimal number of clusters for K-Means, which is visually reinforced by a change in marker size and color.
* **PCA Biplot:** The biplot not only displays the reduced dimensionality data but also overlays feature vectors that indicate the directions and magnitudes of each original feature. This dual representation assists in interpreting the PCA results.
* **Scatterplot Matrix:** The matrix facilitates the exploration of pairwise relationships between the most significant features, allowing users to identify potential correlations and trends across different clusters.

**Conclusion**

The Traffic Accidents Analysis Dashboard offers a powerful, interactive tool for exploring complex accident data. By combining PCA for dimensionality reduction and K-Means clustering for pattern detection, the dashboard enables users to uncover hidden relationships within the dataset. The multi-tab layout ensures that the analysis remains organized and focused, allowing for both high-level overviews and detailed investigations. Ultimately, this project not only demonstrates the utility of advanced data visualization techniques but also provides practical insights that could inform strategies to enhance road safety and reduce accident rates.